

Robust 3D Multi-Camera Tracking from 2D Mono-Camera Tracks by Bayesian Association

Raúl Mohedano, Narciso García
Universidad Politécnica de Madrid, Spain

Abstract— Visual tracking of people is essential automatic scene understanding and surveillance of areas of interest. Monocular 2D tracking has been largely studied, but it usually provides inadequate information for event interpretation, and also proves insufficiently robust, due to view-point limitations (occlusions, etc.). In this paper, we present a light but automatic and robust 3D tracking method using multiple calibrated cameras. It is based on *off-the-shelf* 2D tracking systems running independently in each camera of the system, combined using Bayesian association of the monocular tracks. The proposed system shows excellent results even in challenging situations, proving itself able to automatically boost and recover from possible errors.

I. INTRODUCTION

Robust positioning and tracking of multiple people is a very active research topic due to its applicability to surveillance and security. Although monocular systems have been largely studied in the literature [1] and they are useful in a number of situations, they prove themselves insufficient in complex scenarios, also providing limited information for scene understanding. Both aspects can be successfully addressed by means of 3D tracking from multiple overlapped cameras.

There are a great number of methods for 3D tracking using multiple cameras, differing in several of aspects. Some combine the different views using planar homographies [2], and others directly in the 3D world using calibrated cameras (inferring more information from the scene). In addition, it is possible to use 2D tracking systems (running independently in each camera), combining the resulting 2D tracks, but also to gather all the multi-camera information in a common frame without making partial decisions [3]. The latter can produce good results, but prevent the cameras from performing complex local operations for data reduction.

In this paper, we propose a 3D tracking system based on multiple overlapped cameras that rely on generic 2D tracking systems acting independently in each of the cameras of the system. The system infers which 2D tracks from different cameras correspond to a certain actual 3D object using a Bayesian data association method, allowing a posterior Bayesian tracking. This Bayesian processing allows the utilization of low-complexity 2D tracking modules, making the system both flexible and powerful.

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2007-67764 (SmartVision) and by the Comunidad de Madrid under project S-0505/TIC-0223 (Pro-Multidis). Also, R. Mohedano wishes to thank the Comunidad de Madrid for a personal research grant.

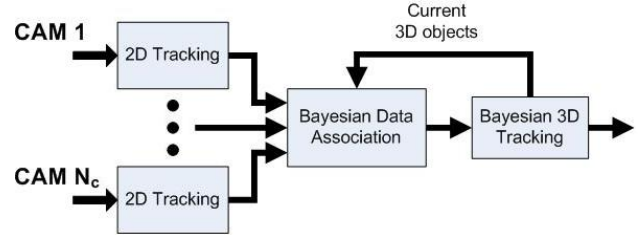


Fig. 1. Block diagram of the proposed system: individual, monocular 2D tracking systems, Bayesian association of 2D tracks to actual 3D objects, and final 3D tracking by combining previously associated data.

II. 3D TRACKING SYSTEM DESCRIPTION

The proposed 3D tracking system, depicted in Figure 1, consists of two different levels: data association, where the correspondence between 2D tracks and actual 3D objects is analyzed, and multi-view info fusion, where 3D parameters of the moving objects are estimated using the previously associated data. In both steps, position calculations are performed using ‘head positions’, i.e. the topmost point of 2D and 3D objects, as they presumably are highly stable under common occlusion situations.

A. Bayesian Data Association

Let us suppose that the system has been successfully tracking N_{OBJ} three-dimensional objects up to time step $t-1$. It implies that both their position-velocity and their appearance (normalized color histogram) are known. The Bayesian data association module aims to select the most likely set of 2D tracks given the currently 3D tracked objects. The sets of tracks supporting the 3D objects will be composed of, at most, one 2D track per camera (per 3D object).

The probability of a given set of 2D tracks can be written as

$$\begin{aligned}
 p(t_{c_1}^{o_1}, \dots, t_{c_N}^{o_1}, \dots, t_{c_1}^{o_M}, \dots, t_{c_N}^{o_M} | o_1, \dots, o_M) &= \\
 &= \prod_{i=1}^M p(t_{c_1}^{o_i}, \dots, t_{c_N}^{o_i} | o_i) = \prod_{i=1}^M \prod_{j=1}^N p(t_{c_j}^{o_i} | o_i),
 \end{aligned}$$

where $t_{c_j}^{o_i}$ represents the 2D track assumed for the 3D object o_i in the camera c_j , N is the number of cameras of the system and M the number of objects. In addition, if we separate spatial and appearance contributions, we can write

$$p(t_{c_j}^{o_i} | o_i) = p_{AP}(h_{c_j}^{o_i} | h_i) p_{SP}(x_{c_j}^{o_i} | x_i)$$

where $h_{c_j}^{o_i}$ and h_i represent the normalized color histogram of $t_{c_j}^{o_i}$ and the object o_i respectively, $x_{c_j}^{o_i}$ is the ‘head position’ of the 2D track, and x_i is the ‘3D head position’ of the 3D object o_i .

The appearance observation model is based on the Bhattacharyya distance (D_B) between $h_{c_j}^{o_i}$ and h_i , and it follows an exponential distribution: $p_{AP} = \lambda \exp(-\lambda D_B)$. The exact value of λ is not critical within a great range, and it is selected empirically. The spatial factor, however, is assumed to be Gaussian, centered at the projection of the 'head position' x_i of the object o_i , and with a covariance related to the distance between the object and the camera.

As explained, we select the most likely combination of 2D tracks (if it is likely enough). The rest of the 2D tracks, then, are either errors, or views of actual 3D objects that are not yet 'registered' in the system. Thus, it is necessary to check if certain set of the remaining 2D tracks is coherent enough to correspond to a real 3D object. For that purpose, for each set of 2D tracks, we create a 'potential 3D object', whose '3D head position' is inferred from the 'head positions' of the 2D tracks using triangulation, and whose appearance is described by the mean of the histograms of the 2D tracks of the set.

B. Bayesian 3D Tracking using Particle Filters

Once correspondence between 2D tracks and 3D objects has been established, 3D object position and appearance must be updated accordingly. We propose Bayesian tracking techniques for the estimation of the object parameters. In particular, we use Particle Filters [4].

We model 3D objects as cylinders, as they are a reasonable approximation of the shape of a human. The hidden variable considered is composed of x - y position and velocity, height and radius. The observation model considered is similar to the model used in the data association step. The system sets an independent Particle Filter for each 3D object, as it has been observed that, in this case, the joint consideration of all the moving objects [5] does not yield significant improvements.

III. EXPERIMENTAL RESULTS

The proposed scheme has been tested in different indoor and outdoor environments, with different illumination conditions, monitored using from two to six cameras. As for monocular 2D tracking, a simple blob graph matching method has been used [6]. The performed tests show the highly satisfying performance of the system, and prove its outstanding capability of handling complex situations and transitory failures of the 2D tracking modules.



Fig. 2. Two examples of the performance of the system (only one of the views presented). Left, precise height measurement (man climbing a stepladder). Right, four interacting people tracking.

As an example, we present here a set of tests performed in

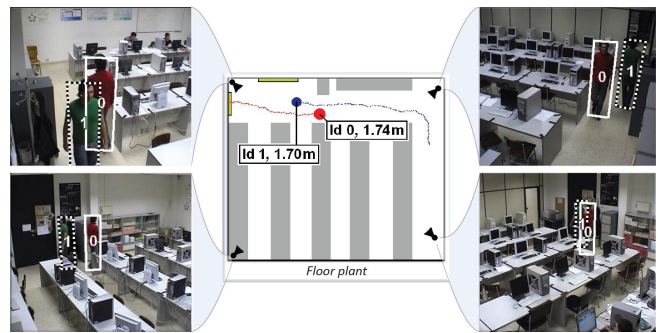


Fig. 3. 3D tracking results in a typical indoor scene, with 2 people interaction (left and right, 3D tracking results re-projected in each of the 4 views, and middle, synthetic bird-eye view of the scene).

an office, as a representative complex indoor scenario. This environment, a room with two entrance doors and furniture, has been monitored using four calibrated cameras. Different situations have been evaluated: different numbers of interacting people, entering and leaving the scene, changing direction, and interacting closely in different ways. Figures 2 and 3 show the 3D positioning quality of the complete system, proving its performance in different situations.

IV. CONCLUSIONS

We have presented an accurate and robust multi-camera 3D tracking method based on a probabilistic combination of the tracking info acquired independently by each camera of the system. This allows the use of *off-the-shelf* 2D tracking modules, widely available and computationally inexpensive but prone to errors, to build a complete 3D tracking system with high accuracy and reliability. Different tests performed on the systems show its capability of automatically handle potential 2D tracking errors.

The system achieves an accurate 3D positioning and tracking by combining multi-view info according to geometric and appearance consistency. To provide the system with higher robustness in complex scenarios (with multiple occluding objects), geometrical calculations are performed using head positions, less variant, instead of centroids. In addition, the proposed system proves able to automatically handle the entrance of new objects of the scene, unlike many more complex multi-camera tracking systems.

REFERENCES

- [1] A. Yilmaz, O. Javed and M. Shah, "Object tracking: a survey", *ACM Computing Surveys (CSUR)*, vol. 38, no. 4, 2006.
- [2] R.I. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", 2nd ed., Cambridge University Press, Cambridge, 2004.
- [3] S.M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 505-519, 2009
- [4] M.S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on Particle Filters for online nonlinear/non-Gaussian Bayesian tracking", *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174-188, 2002.
- [5] Z. Khan, T. Balch and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1805-1819, 2005.
- [6] C. Gomilla and F. Meyer, "Graph-based object tracking", *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 41-44, 2003.